

Table 1

Protein Seq ID No:	Nucleotide Seq ID No:	Clone ID	Library	Fragments
1	19	155179	THP1PLB02	155179H1 (THP1PLB02), 155179X307D2 (THP1PLB02), 121411X24 (BRSTTUT01), 961990X21 (BRSTTUT03), 961990X17 (BRSTTUT03), 478199X15 (MMLR2DT01), 692774X19 (LUNGTTUT02), 034109F1 (THP1NOB01), 2754059H1 (THP1AZS08)
2	20	2415780	HNT3AZT01	2415780H1 (HNT3AZT01), 1443076F6 (THYRN0T03), 1753156F6 (LIVRTUT01), 989015H1 (LVEEN0T03), 1922409R6 (BRSTTUT01)
3	21	2879274	UTRSTUT05	2879274H1 (UTRSTUT05), 3537571H1 (SEMVN0T04), 2879274H1 (UTRSTUT05), 2767241H1 (COLAN0T02), 1479540F1 (CORPN0T02), 1650591F6 (PROSTUT09), 1650591T6 (PROSTUT09), 1264516R1 (SYNORAT05), 1438281F1 (PANCON0T08)
4	22	358050	PROSN0T01	041451R6 (TBLYN0T01), 358050H1 (PROSN0T01), 1288739F6 (BRAIN0T11), 1338092F6 (COLNN0T13), 1338092T6 (COLNN0T13), 1817810F6 (PROSN0T20), 3049061H1 (LUNGNOT25), 3217540H1 (TESTNOT07), 3224582H2 (UTPRSNON03)
5	23	700745	SYNORAT03	700745H1 (SYNORAT03), 700745R6 (SYNORAT03)
6	24	2026480	KERAN0T02	1288279F6 (BRAIN0T11), 1798769H1 (COLNN0T27), 1984648T6 (LUNGAST01), 2026480H1 (KERAN0T02), 3577373F6 (BRONN0T01), 4049569T6 (SINTNOT18), SAYA00492F1
7	26	2132401	OVARN0T03	014071R6 (THP1PLB01), 2669596F6 (ESOGTUT02), 4511344H1 (EPIMN0T01), SAJA01969F1, SAJA00384R1, SAJA00561F1
8	26	2568875	HIP0AZT01	826204R1 (PROSN0T06), 826204X144F1 (PROSN0T06), 826204X48 (PROSN0T06), 826204X52 (PROSN0T06), 2568875H1 (HIP0AZT01)
9	27	3408908	PROSTUS08	3408908F6 (PROSTUS08), 3408908H1 (PROSTUS08), SBWA03204V1
10	28	3772696	BRSTNOT25	3772696F6 (BRSTNOT25), 3772696H1 (BRSTNOT25), 3772696T6 (BRSTNOT25), SXBA00825V1, SXBA00411V1

Table 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
11	29	5388674	BRAINOT19	2260285X320D4 (UTRSNOT02), 2287941X301D1 (BRAINON01), 2289534R6 (BRAINON01), 3521165H1 (LUNGON03), 4315221H1 (BRAFN0T01), 5082079H1 (LNODNOT11), 5388674H1 (BRAINOT19)
12	30	1873102	LEUKNOT02	092669F1 (HYPONOB01), 225519F1 (PANCNOT01), 225519R1 (PANCNOT01), 390991H1 (TMLR2DT01), 1737263F6 (COLNNOT22), 1737263T6 (COLNNOT22), 1873102H1 (LEUKNOT02), 1932133F6 (COLNNOT16), 3590995H1 (293TF5T01), 3712151H1 (PENCNOT09), 4285941H1 (LIVRDIR01), 4339405H1 (BRAUNOT02)
13	31	1920734	BRSTTUT01	991651H1 (COLNNOT11), 1920734H1 (BRSTTUT01), 1920734R6 (BRSTTUT01), 1920734T6 (BRSTTUT01), 2739282F6 (OVARNOT09), 3765480H1 (BRSTNOT24)
14	32	2396858	THPIAZT01	1439237F1 (PANCNOT08), 1722122F6 (BLADNOT06), 1908978F6 (CONNTUT01), 2396858H1 (THPIAZT01), 2396858X301V1 (THPIAZT01), 2396858X305D1 (THPIAZT01), 2461972F6 (THYRN0T08)
15	33	2634725	COLNTUT15	1875442H1 (LEUKNOT02), 2634725F6 (COLNTUT15), 2634725H1 (COLNTUT15), 2920995T6 (SININOT04), 4875374H1 (COLDNOT01)
16	34	2643110	LUNGTTUT08	881275H1 (THYRN0T02), 1273883X302D2 (TESTTUT02), 1273883X304D2 (TESTTUT02), 1918031R6 (PROSN0T06), 2171263F6 (ENDCNOT03), 2453207F6 (ENDANOT01), 2453207T6 (ENDANOT01), 2643110H1 (LUNGTTUT08), 2753878H1 (THPIAZS08)
17	35	2701396	OVARTTUT10	2701396H1 (OVARTTUT10), 2867440T6 (KIDNNOT20), SBLA01199F1, SBLA03620F1, SBLA02714F1
18	36	3134404	SMCCNOT01	3134404H1 (SMCCNOT01), 4161423F6 (BRSTNOT32), 4384476H1 (BRAVTTT02), SCAA06693V1

Table 2

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods
1	762	S174 S190 T225 S341 S417 T526 T681 S717 S732 S94 T166 T227 S329 T333 T365 T368 T402 T436 T478 S589 T692 Y342	N16 N209 N392 N679	Cell attachment sequence: R156-D158 Serine protease trypsin family active sites: C548-C564, V559-C564, V654-T660, D706-S717, C708-S718, W733-I756 Trypsin motif: V522-I756 Chymotrypsin serine protease family: G549-C564, D614-A628, V705-S717 Low-density lipoprotein receptor: D358-C395, G369-E390, C371-E390, A397-C432, K406-E427, P433-K468, G442-E463, V472-C511 Kringle motif: C548-Y565, D705-S718 Developmental CUB domain: C121-F238, C247-Y351	Epithin (membrane bound serine protease) [Mus musculus] g4104970  Enterokinase [Homo sapiens] g746413 (21% identity)	MOTIFS BLOCKS PRINTS PFAM BLAST

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods
2	335	T67 S209 T225 S232 T45 T196 T245 T321 T332 Y195		Glycoprotease family: V4-V18, T38-L66, D70-I114, R94-I114, T127-I139, Y146-I170, C252-E271 Glycoprotease motif: P2-R308 O-sialoglycoprotein endopeptidase: L5-V18, I74-I114, T127-I139, F149-I170, A257-N266	O-sialoglycoprotease [Rattus norvegicus] g5360708  O-sialoglycoprotein endopeptidase [Methanococcus jannaschii] g2826367 (44% identity)	MOTIFS BLOCKS PRINTS PFAM BLAST
3	327	S152 S166 T175 S285 S292 S48 T73	N150 N191 N198 N263	Signal peptide: M1-A26	Prolylcarboxy- peptidase [Homo sapiens] g431321 (33% identity)	MOTIFS SPScan BLAST
4	471	S44 S468 S26 S47 S64 T82 S117 T244 T280 S445 S40 T69 S145 T307 T405 Y106 Y223		F-box domain: P10-H56 Signal peptide: M1-L33	F-box protein sequence (GeneSeq Y02253)	MOTIFS PFAM SPScan BLAST
5	60	S15		Signal peptide: M1-G20 Tissue inhibitor of metalloproteinases signature: G17-C46	TIMP-3 (Tissue inhibitor of metalloproteinases-3) [Homo sapiens] g1215682	BLAST MOTIFS SPSCAN BLOCKS HMM ProfileScan

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods
6	399	S24 S45 S30 S163 T211 S244 T321			PINT domain protein (Proteasomal subunit) [Plasmodium falciparum] g3845132	BLAST MOTIFS
7	106	S6 S20 T32 T102 S63 S69 T74		Ubiquitin signature: V39-V90		MOTIFS BLOCKS
8	267	S2 T72 T89 S211 S236 S12 S111	N260	Trypsin serine protease active site: L51-C56 Trypsin serine protease signature: T15-V235 Signal peptide: M1-Q61	Prostasin (serine protease) [Homo sapiens] g1143194	BLAST MOTIFS PFAM BLOCKS PRINTS SPSCAN ProfileScan
9	123	S44 S80 T58 S75 S103		Signal peptide: M1-G24 WAP domain: K30-P72, K77-P120	Secretory leukocyte protease inhibitor [Mus musculus] g1763263	BLAST MOTIFS SPSCAN PFAM PRINTS HMM ProfileScan

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods
10	513	T24 S57 T193 S249 T311 T75 T88 T112 T290 T384 T385 S422	N55 N110 N200 N452 N470 N508	Signal peptide: M1-P19 Matrxin signature: Y36-T202 Neutral zinc metalloproteases Zn-binding region: V213-L222 Hemopexin domain: F285-C465	Matrix metalloproteinase [Gallus gallus] g3511149	BLAST MOTIFS SPScan PFAM BLOCKS PRINTS HMM ProfileScan
11	326	S191 T140 S158 S196 S269 S296 T26 T41 S104 S214	N34	Neutral zinc metalloproteases Zn-binding region: T217-G227		MOTIFS BLOCKS ProfileScan
10	823	T6 S350 S168 T277 S353 S381 S398 T407 S415 S479 S524 S531 S566 T641 T97 T146 T194 T271 T277 T331 S394 T435 T658 T727 T753 T806	N111 N213 N329 N421 N596	Ubiquitin carboxyl-terminal hydrolases family 2: G197-L214, Y295-L304, V355-C369, L741-A765, Y742-Y760, K790-N811 Ubiquitin carboxyl-terminal hydrolase family: G197-L214, Y742-V801	Ubiquitin-specific protease UBP41 [Mus musculus] g3386552	MOTIFS BLAST PFAM BLIMPS
13	404	S58 S68 T107 S164 T177 T208 S284 T14 S68 T341	N339		Similar to zinc metalloprotease [C. elegans] g2804437	MOTIFS BLAST

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods
14	703	S20 S68 T120 T135 S331 T383 S562 S606 S607 S631 S674 S698 T31 S95 S115 S173 S355 S490 S562 S650	N318 N434 N445 N670	E1 ubiquitin activating enzyme: K352-H442	E1-like protein (ubiquitin activating enzyme) [Pichia pastoris] g4262402	MOTIFS BLAST BLIMPS
15	145	T36 S100 S115 T47	N34	Protease serine hydrolase precursor signal zymogen glycoprotein multigene family: L16-Q64, G87-K140 Trypsin: L25-Q64, S84-N142	Matriptase (serine protease) [Homo sapiens] g5359675, g6002714 Epithin (membrane bound serine protease) [Mus musculus] g4104970	MOTIFS BLAST BLIMPS
16	518	S74 T252 S151 T169 T245 S312 S361 T419 S462 S502 S16 S70 S98 S133 T301 S331 S428 T516 Y334	N234	Dipeptidyl peptidase IV: H255-L305, E326-Q352, E379-P411	Dipeptidyl peptidase IV [Stenotrophomonas maltophilia] g1753197	MOTIFS BLAST BLIMPS

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods
17	476	S25 T183 S203 S324 S358 T398 S441 S457 T471 S472 S25 S345 T398 S402 T420 Y34 Y447	N38 N175 N314 N360 N455	Eukaryotic thiol proteases active site: F431-I450 Cysteine protease: C240-W467 Eukaryotic thiol protease active site: K237-F246, R281-I289, T410-G419, F431-Y447	Similar to cysteine protease [ <i>C. elegans</i> ] g3876422 Cathepsin B [ <i>Triticum aestivum</i> ] g21693	MOTIFS BLAST BLIMPS PFAM BLIMPS
18	229	S96 S219 S77 S78	N11 N105 N125	Signal peptide: M1-C25 Zn metalloprotease: S63-L210	Zinc metalloprotease ADAMTS7 [ <i>Homo sapiens</i> ] g5923788	MOTIFS BLAST HMM BLIMPS



Table 3

Nucleotide Seq ID NO:	Useful Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
19	322-366	Gastrointestinal (0.33) Reproductive (0.28) Hematopoietic/Immune (0.28)	Cell Proliferation (0.66) Inflammation and Immune Response (0.31)	
21	499-543	Reproductive (0.21) Hematopoietic/Immune (0.20) Nervous (0.19)	Cell Proliferation (0.59) Inflammation and Immune Response (0.43)	
21	1082-1126	Reproductive (0.30) Nervous (0.18) Gastrointestinal (0.12)	Cell Proliferation (0.61) Inflammation and Immune Response (0.34)	
22	305-478 1847-1891	Reproductive (0.360) Nervous (0.220) Cardiovascular (0.100)	Cell Proliferation (0.560) Inflammation and Immune Response (0.200)	PBLUESCRIPT
23	146-190	Reproductive (0.500) Developmental (0.250) Musculoskeletal (0.250)	Cell Proliferation (0.250) Inflammation and Immune Response (0.250)	PSPORT1
24	433-477	Reproductive (0.250) Gastrointestinal (0.155) Hematopoietic/Immune (0.155) Nervous (0.155)	Cell Proliferation (0.667) Inflammation and Immune Response (0.274)	PSPORT1
25	56-100 440-484	Gastrointestinal (0.207) Reproductive (0.207) Cardiovascular (0.103) Hematopoietic/Immune (0.103) Musculoskeletal (0.103) Nervous (0.103)	Cell Proliferation (0.589) Inflammation and Immune Response (0.448)	PSPORT1

Table 3 (cont.)

Nucleotide SEQ ID NO:	Useful Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
26	704-748 1001-1045	Reproductive (0.500) Cardiovascular (0.125) Gastrointestinal (0.125) Hematopoietic/Immune (0.125) Nervous (0.125)	Cell Proliferation (0.375) Inflammation and Immune Response (0.250)	PSPORT1
27	189-233 327-371	Reproductive (0.667) Dermatologic (0.333)	Cell Proliferation (0.667) Inflammation and Immune Response (0.333)	PT7T3
28	168-212 1227-1271	Reproductive (1.000)	Inflammation and Immune Response (1.000)	PINCY
29	226-270	Reproductive (0.258) Nervous (0.194) Hematopoietic/Immune (0.172)	Cell Proliferation (0.591) Inflammation and Immune Response (0.376)	PINCY
30	649-693	Reproductive (0.235) Hematopoietic/Immune (0.163) Nervous (0.153)	Cancer (0.418) Inflammation (0.276) Cell Proliferation (0.163)	PINCY
31	379-423	Reproductive (0.348) Nervous (0.217) Cardiovascular (0.174)	Cancer (0.435) Inflammation (0.130) Cell Proliferation (0.087) Trauma (0.087)	PSPORT1
32	704-748	Reproductive (0.262) Hematopoietic/Immune (0.167) Nervous (0.143)	Cancer (0.500) Inflammation (0.262) Cell Proliferation (0.214)	PINCY

Table 3 (cont.)

Nucleotide Seq ID NO:	Useful Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
33	433-477	Gastrointestinal (0.365) Reproductive (0.288) Hematopoietic/Immune (0.115)	Cancer (0.538) Inflammation (0.250) Trauma (0.115)	pINCY
34	1398-1442	Reproductive (0.200) Cardiovascular (0.150) Gastrointestinal (0.150) Nervous (0.150)	Cancer (0.350) Cell Proliferation (0.300) Inflammation (0.150)	pINCY
35	755-801	Urologic (0.500) Gastrointestinal (0.167) Nervous (0.167) Reproductive (0.167)	Cancer (0.667) Trauma (0.333)	pINCY
36	447-491	Reproductive (0.375) Cardiovascular (0.250) Developmental (0.125) Nervous (0.125) Urologic (0.125)	Cell Proliferation (0.500) Other (0.250) Inflammation (0.125) Trauma (0.125)	pINCY

# Table 4

SEQ ID NO:	Library	Library Comments
19	THP1PLEB02	Library was constructed by reamplification of a human promonocyte line library, which was made using RNA isolated from THP-1 cells cultured for 48 hours with 100 ng/ml phorbol ester (PMA), followed by a 4-hour culture in media containing 1 ug/ml LPS. THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year-old male with acute monocytic leukemia.
20	HNT3AZT01	Library was constructed using RNA isolated from the hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor). Cells were treated for three days with 0.35 micromolar 5-aza-2'-deoxycytidine (AZ).
21	UTRSTUT05	Library was constructed using RNA isolated from uterine tumor tissue removed from a 41-year-old Caucasian female during a vaginal hysterectomy with dilation and curettage. Pathology indicated uterine leiomyoma. The endometrium was secretory and contained fragments of endometrial polyps. Benign endo- and ectocervical mucosa were identified in the endocervix. Patient history included a ventral hernia and a benign ovarian neoplasm.
22	PROSNOT01	Library was constructed using RNA isolated from the prostate tissue of a 78-year-old Caucasian male, who died from leukemia. Patient history included skin cancer, emphysema, and asthma. Previous surgeries included a cholecystectomy.
23	SYNORAT03	Library was constructed using RNA isolated from the wrist synovial membrane tissue of a 56-year-old female with rheumatoid arthritis.
24	KERANOT02	Library was constructed using RNA isolated from epidermal breast keratinocytes (NHEK). NHEK (Clontech #CC-2501) is a human breast keratinocyte cell line derived from a 30-year-old black female during breast-reduction surgery.
25	OVARNOT03	Library was constructed using RNA isolated from ovarian tissue removed from a 43-year-old Caucasian female during removal of the fallopian tubes and ovaries. Pathology for the associated tumor tissue indicated grade 2 mucinous cystadenocarcinoma. Patient history included mitral valve disorder, pneumonia, and viral hepatitis. Family history included atherosclerotic coronary artery disease, pancreatic cancer, stress reaction, cerebrovascular disease, breast cancer, and uterine cancer.

Table 4 (cont.)

SEQ ID NO:	Library	Library Comments
26	H1POAZT01	Library was constructed from RNA isolated from diseased hippocampus tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.
27	PROSTUS08	Library was constructed using 2.36 million clones from a prostate tumor library and was subjected to one round of subtractive hybridization with 448,000 clones from a control prostate library. The starting library for subtraction was constructed using RNA isolated from a prostate tumor removed from a 59-year-old Caucasian male during a radical prostatectomy with regional lymph node excision. Pathology indicated adenocarcinoma (Gleason grade 3+3). Adenofibromatous hyperplasia was present. The patient presented with elevated prostate-specific antigen (PSA). Patient history included colon diverticuli, asbestosis, and thrombophlebitis. Family history included multiple myeloma, hyperlipidemia, and rheumatoid arthritis. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., Nucleic Acids Res. (1991) 19:1954 and Bernaldo et al., Genome Research (1996) 6:791.
28	BRSTNOT25	Library was constructed using RNA isolated from breast tissue removed from a 35-year-old Caucasian female during a bilateral reduction mammoplasty. Family history included uterine cancer, hyperlipidemia, benign hypertension, acute myocardial infarction, cerebrovascular disease, atherosclerotic coronary artery disease, and type II diabetes.
29	BRAINOT19	Library was constructed using RNA isolated from diseased brain tissue removed from the left frontal lobe of a 27-year-old Caucasian male during a brain lobectomy. Pathology indicated a focal deep white matter lesion, characterized by marked gliosis, calcifications, and hemosiderin-laden macrophages, consistent with a remote perinatal injury. This tissue also showed mild to moderate generalized gliosis, predominantly subpial and subcortical, consistent with chronic seizure disorder. The left temporal lobe, including the mesial temporal structures, showed focal, marked pyramidal cell loss and gliosis in hippocampal sector CA1, consistent with mesial temporal sclerosis. GFAP was positive for astrocytes. The patient presented with intractable epilepsy, focal epilepsy, hemiplegia, and an unspecified brain injury. Patient history included cerebral palsy, abnormality of gait, and depressive disorder. Family history included brain cancer.

Table 4 (cont.)

SEQ ID NO:	Library	Library Comments
30	LEUKNOT02	Library was constructed using RNA isolated from white blood cells of a 45-year-old female with blood type O+. The donor tested positive for cytomegalovirus (CMV).
31	BRSTTUT01	Library was constructed using RNA isolated from breast tumor tissue removed from a 55-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated invasive grade 4 mammary adenocarcinoma of mixed lobular and ductal type, extensively involving the left breast. The tumor was identified in the deep dermis near the lactiferous ducts with extracapsular extension. Proliferative fibrocystic changes were characterized by apocrine metaplasia, sclerosing adenosis, cyst formation, and ductal hyperplasia without atypia. Patient history included atrial tachycardia, blood in the stool, and a benign breast neoplasm. Family history included benign hypertension, atherosclerotic coronary artery disease, cerebrovascular disease, and depressive disorder.
32	THP1AZT01	Library was constructed using polyA RNA isolated from THP-1 promonocyte cells treated for three days with 0.8 micromolar 5-aza-2'-deoxycytidine. THP-1 (ATCC TIB 202) is a human promonocyte line derived from blood of a 1-year-old Caucasian male with acute monocytic leukemia.
33	COLNTUT15	Library was constructed using RNA isolated from colon tumor tissue obtained from a 64-year-old Caucasian female during a right hemicolectomy with ileostomy and bilateral salpingo-oophorectomy (removal of the fallopian tubes and ovaries). Pathology indicated an invasive grade 3 adenocarcinoma. Patient history included hypothyroidism, depression, and anemia. Family history included colon cancer and uterine cancer.

Table 4 (cont.)

SEQ ID NO:	Library	Library Comments
34	LUNGSTUT08	Library was constructed using RNA isolated from lung tumor tissue removed from a 63-year-old Caucasian male during a right upper lobectomy with fiberoptic bronchoscopy. Pathology indicated a grade 3 adenocarcinoma. Patient history included atherosclerotic coronary artery disease, an acute myocardial infarction, rectal cancer, an asymptomatic abdominal aortic aneurysm, and cardiac dysrhythmia. Family history included congestive heart failure, stomach cancer, and lung cancer, type II diabetes, atherosclerotic coronary artery disease, and an acute myocardial infarction.
35	OVARUTUT10	Library was constructed using RNA isolated from ovarian tumor tissue removed from the left ovary of a 58-year-old Caucasian female during a total abdominal hysterectomy, removal of a solitary ovary, and repair of inguinal hernia. Pathology indicated a metastatic grade 3 adenocarcinoma of colonic origin, forming a partially cystic and necrotic tumor mass in the left ovary, and an adenocarcinoma of colonic origin, forming a nodule in the left mesovarium. A single intramural leiomyoma was identified in the myometrium. The cervix showed mild chronic cystic cervicitis. Patient history included benign hypertension, follicular cyst of the ovary, colon cancer, benign colon neoplasm, and osteoarthritis. Family history included emphysema, myocardial infarction, atherosclerotic coronary artery disease, benign hypertension, and hyperlipidemia.
36	SMCCNOT01	Library was constructed using RNA isolated from smooth muscle cells removed from the coronary artery of a 3-year-old Caucasian male.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	<i>ESTs</i> : Probability value= 1.0E-8 or less <i>Full Length sequences</i> : Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, and tsearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	<i>ESTs</i> : fasta E value=1.06E-6 <i>Assembled ESTs</i> : fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less <i>Full Length sequences</i> : fastx score=100 or greater
BLIMPS	A BLOCKS IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families



# Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfilesScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Philis Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPSscan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. <u>supra</u> : Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	